Enhancing Model-Based Systems Engineering with Model Informativity Analysis

Research Progress Report

June 2017

Inbar Zigdon Batan, Yaniv Mordecai, Niva Wengrowitz,
Avigdor Zonnenshein, and Dov Dori

Technion – Israel Institute of Technology, Haifa 32000, Israel

EXECUTIVE ABSTRACT

Model Based System Engineering (MBSE) is the application of formal qualitative and quantitative modeling to support the system definition, design, development, deployment, and operation, i.e., the entire system's lifecycle. MBSE relies on modeling languages, such as Object-Process Methodology (OPM) or Systems Modeling Language (SysML). Modeling is done using software tools, such as OPM's free modeling tool, OPCAT, IBM's Rational Rhapsody, or MathWorks' Matlab-Simulink. The tools help system analysts model the system, its function, structure, and behavior. However, most organizations developing complex systems still do not use MBSE; rather, they base their systems engineering process on textual documentation and occasional diagrams.

Models of complex engineered systems capture and provide information and knowledge about the systems and their components, and may therefore become very big and complex themselves. It is difficult to measure the amount and quality of information in the model, to identify information gaps, or to assess the model's descriptiveness of the system or the sufficiency of the information about the system. Currently no tool measures, monitors, and analyzes different informational model aspects, and allows the systems engineer to determine that a model is sufficiently descriptive, contributing to the understanding of the system, or missing critical information. Hence, it is not possible to assess the value of system models in terms of the information they convey and their contribution to designing, developing, testing, deploying, or operating the system.

We propose a structure of metrics, Model Informativity Level (MIL), that measure and aggregate the amount, quality, and utility of the various information aspects in the model. MIL provides an estimate for the model's expressive power, descriptiveness, and usability. It also helps identify aspects with insufficient informativity and consequently allocate engineering and modeling efforts. We provide metric visualizations that mediate the information to model stakeholders and systems engineers. In addition, we explore MIL evolution along the system modeling and development stages by applying algorithms for calculating MIL for OPM models constructed by individuals and teams in both academia and industry.

Our findings show that when modeling a system in the MBSE approach using OPM, the model contains the system environment, requirements, main flow and alternative flows of the system. The MIL evaluation also shows that the model informativity increases along the modeling processes, and there are almost no model parts which are remodeled. In addition, by using MIL outcomes, the system engineer receives feedback of the system informativity, and can see where there are gaps in information.

This research was motivated by the assertion that quantitative assessment of model utility and model-based system development trend analysis will help drive the adoption of a MBSE approach along the lifecycle of the system. As part of this research, we have established an open-access repository of MBSE case studies, including OPM models and their informativity scorecards.

1. INTRODUCTION

A system begins with an idea that has to be turned into reality. One good way for designers to accomplish this is to find a clear way to explain how theory meets reality. The system designer can build a model to explain the theory and the problem, to clarify the concept and the system's structure and behavior, represent the environmental constraints, and elaborate the solution design. Simple drawings of phenomena or mechanisms on whiteboards or in textbooks are all models, but they do not meet these requirements. The formality of models is critical for encoding, verification, consistency checking, reproduction, and comparison with other models.

Conceptual models of system are required for understanding the system and simplifying its complexity. *A conceptual model is a formal model of a system which expresses its architecture by depicting its structure and behavior to a level of detail that is sufficient for its subsequent detailed design and eventual materialization* (Dori, 2016). Model-Based Systems Engineering (MBSE) is *the formalized application of modeling to support system requirements, design, analysis, verification and validation activities beginning in the conceptual design phase and continuing throughout development and later life cycle phases* (INCOSE SE VISION 2020, 2007). MBSE provides a framework in which the system engineering team can be effective and consistent from the onset of any project (Long and Scott, 2011).

Object-Process Methodology (OPM) (Dori, 2002, 2016) is a holistic conceptual modeling paradigm for multidisciplinary, complex and dynamic systems and processes. OPM is publicly available as ISO/PAS 19450 [6] for system and process modeling, and listed as a state-of-the-art conceptual modeling and Model-Based System Engineering (MBSE) methodology and paradigm (Estefan, 2008; Ramos, Ferreira, and Barceló, 2012). OPM provides a standardized underlying formalism for problem and solution modeling and MBSE. OPM captures the functional, structural and procedural aspects of the system in a single notation, expressed by both a graphical modality and a textual modality, showing the connections between all the system's objects and processes at varying levels of detail.

Model informativity is a measure of the quality, amount, and utility of the information about the system that is explicitly or implicitly represented by or inferable from the model (Mordecai and Dori, 2016). The informativity level of conceptual system models was first evaluated quantitatively in (Mordecai, 2016; Mordecai and Dori, 2016). They introduced a theoretical informativity evaluation framework for evaluating the informative utility of conceptual models – Model Informativity Analysis (MIA). MIA is a quantitative, utility-based, prescriptive approach for boosting conceptual models' expressive power and for measuring the value of the information they provide (Mordecai and Dori, 2016).

This research was motivated by the intention to improve the awareness of the benefits of MBSE by quantifying the informative value of conceptual system. In this study, we implemented MIA to calculate model informativity level (MIL) of OPM models, and analyzed the informativity scores of various models built by students, professionals, and experts in academia and industry.

2. METHOD

For the purposes of this research, we have developed the **MIA Tool**. The **MIA Tool** generates various **Model Informativity Reports**, which provide different kind of information about the system, based on its OPM model in OPCAT (OPM's free CASE tool). An OPM model of the Model Informativity Analyzing system is shown in Figure 1. The informativity scores are calculated based on the OPL sentences that are automatically generated by OPCAT for the model. Each sentence is scored according to several informativity enhancing factors (IEFs), which are classified under three categories: 1. specification pattern informativity, 2. uncertainty-induced informativity, and 3. meta-specification informativity.

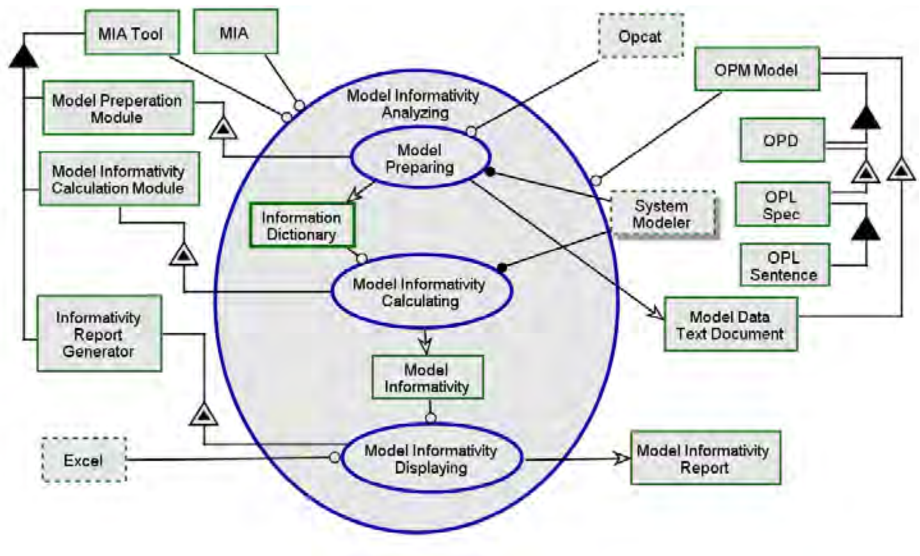

Figure 1 - Model Informativity Analyzing

Each OPL sentence is generated according to a specification pattern. The specification pattern defines how statements and specifications in the model are constructed, based on syntactic and semantic building blocks. OPM's 21 specification patterns are grouped under four categories: thing definition, structural link, procedural link, and precedence link. They vary in importance and significance. Each OPL sentence scored according it spec. pattern's importance, and five more refinements: 1. essence – distinction of things as physical vs informatical; 2. affiliation – pertinence of things to the system or to its environment*;* 3. statefulness – attribution of process-object relations to specific object states; 4. path – attribution of object-process relations to specific flows; and 5. logic – specification of logical conditions AND, OR, and XOR among multiple relations with the same process or object.

Each OPL sentence has some uncertainty associated with it, which affects its informativity. The information theoretic approach defines the amount of information carried in a message (here statement) over a communication channel (here model) as the entropy of the data with respect to its distribution.

Meta-specification is information about the specification, which is not included in the specification itself: rationale (a reasonable justification for the presence of the statement in the model), initiator (client, owner, user, product manager, etc.), priority (the urgency of or preference for developing or implementing the statement), and maturity (the life cycle stage of the statement).

Any OPL sentence in an analyzed model is evaluated on 12 IEFs. Each IEF category has a different contribution to the overall informativity of the OPL sentence, therefore the weight of the categories is different. Furthermore, in each category of IEFs, every IEF has different levels of importance. The weighted INF (WINF) of each sentence in each category is a weighted average of its IEFs. The total WINF (TWINF) of any OPL sentence is a weighted average of the WINFs of the three categories. The informativity level is calculated for each modeling stage, starting from the problem model, through the requirements model, concept model, and solution model.

The implementation is mostly based on the Python 2.7 programing language (van Rossum, 1995). The algorithm input is a model summary text document that the user generates with OPCAT. After generating the document, the user can run a script, which receives necessary details about the model, and generates the following outputs:

A. Model informativity reports:
    a. CSV file with informativity contribution of each OPL statement on each IEF.
    b. CSV file with informativity contribution of each IEF group per stage/ version.
B. Pie Charts which represent the informativity analysis of different IEFs:
    a. MFSP Group informativity distribution pie,
    b. Thing Definition informativity distribution pie,
    c. Structural Links informativity distribution pie,
    d. Procedural – Basic Links informativity distribution pie, and
    e. Procedural – Control Links informativity distribution pie.
C. Pie Charts which represent the system structure:
    a. Objects with structural relation distribution pie, and
    b. Objects – Processes distribution pie.
D. Pie Charts which represent the system behavior:
    a. Objects with behavioral relation distribution pie,
    b. Object essence distribution pie (informatical vs. physical),
    c. Objects affiliation distribution pie (systemic vs. environmental),

d. Processes essence distribution pie (informatical vs. physical), and

e. Processes affiliation distribution pie (systemic vs. environmental).

E. MIL trend graph along the project timeline.

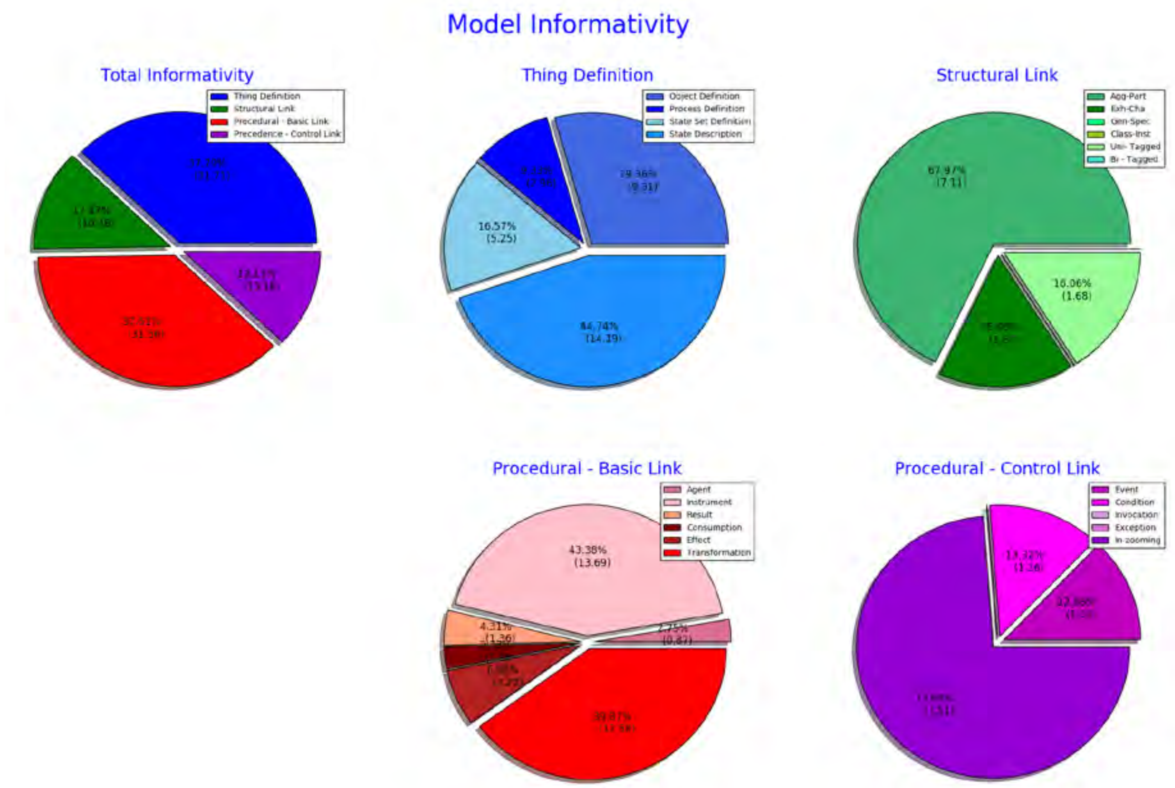

Figure 2 - Model informativity Pie Charts

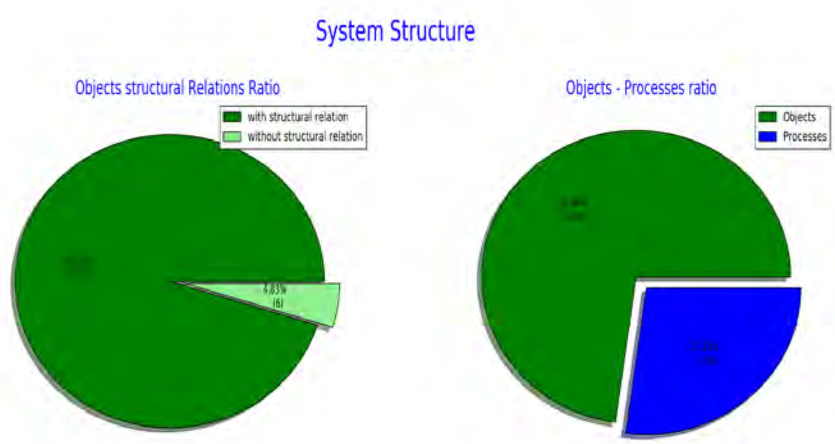

Figure 3 - System Structure Pie Charts

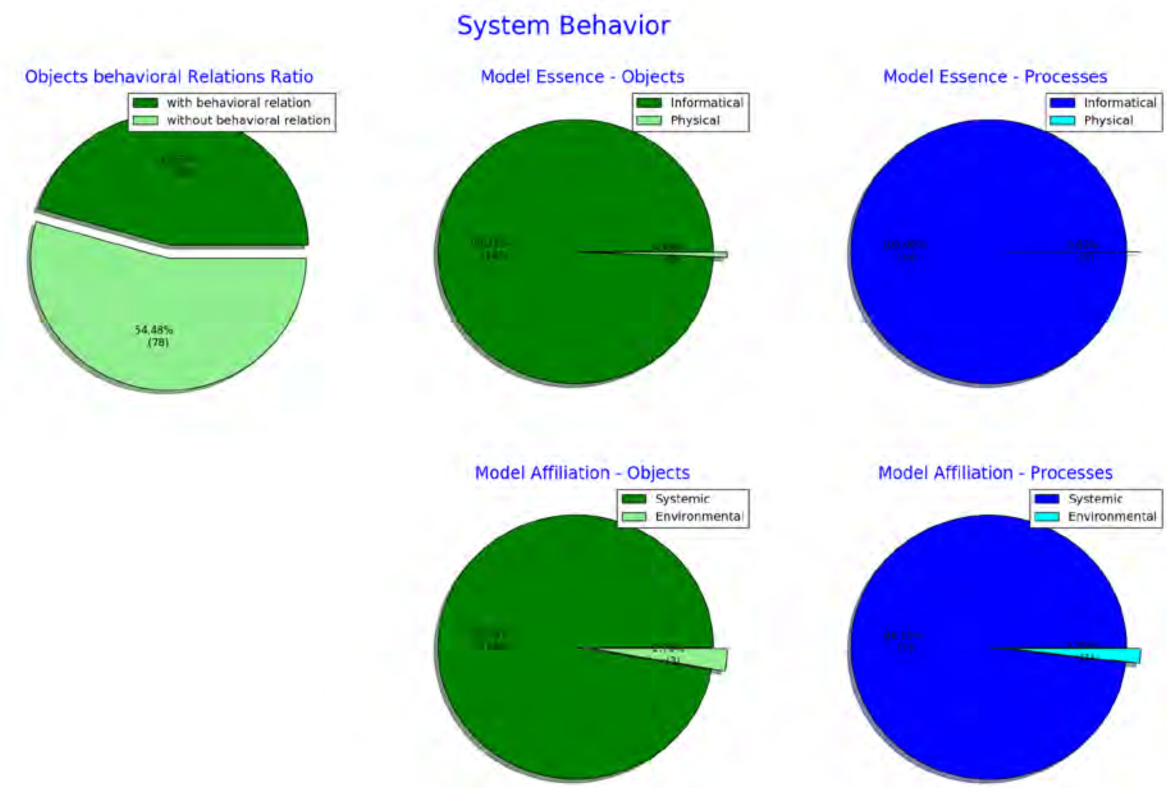

Figure 4 - System Behavior Pie Charts

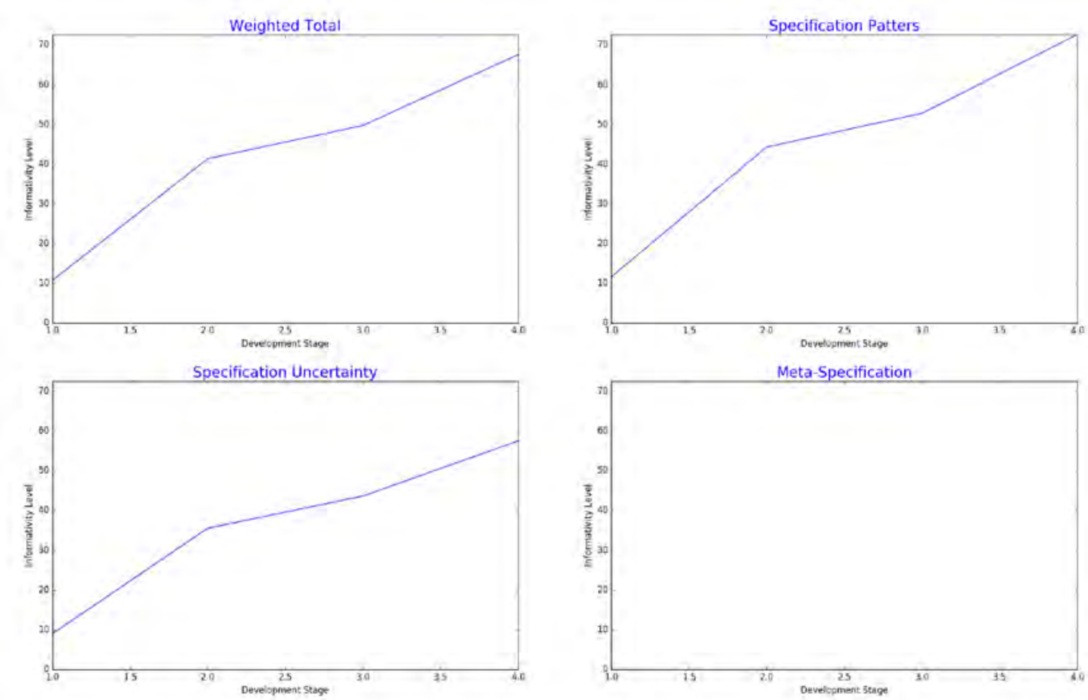

Figure 5 - Model Informativity Trend Graph

3. RESULTS

As part of this study, we have analyzed the informativity of various models. We had four distinct groups of models: 1. versioned complex system models, built by graduate student teams; 2. versioned protocol models, built by modeling experts and student teams; 3. versioned industrial workshop models, built by or for companies for real systems under development or operation; and 4. theoretical models, built by modeling experts for research and teaching purposes.

In order to drive and promote the adoption of MBSE in general and MIA-enhanced MBSE in particular, we built a public online MBSE Repository—a repository of MBSE artifacts available at https://www.dropbox.com/sh/evv2j3kii3kj641/AACOBqWw43ZzG1mGVOJLwjhla?dl=0. The MBSE Repository is divided into the four models groups. Each case study includes the different stages of model developing, reports, and visual graphs. All the models are OPM models.

For each model, we evaluated the trend of the following metrics: 1. Number of OPL sentences in each modeling stage, 2. Specification Patterns Weighted Informativity Figure (WINF) per sentence, 3. Total WINF (TWINF) per sentence, 4. proportion of objects and processes, 5. proportion of systematic and environmental things, 6. proportion of physical and informatical things. We analyzed a total of 42 different models, of which 18 had one phase and 24 had 2-4 phases, as shown in Table 1.

Table 1 - analyzed model groups with modeling stages

Models group	Sample Size	1-modeling stage	2-modeling stage	3-modeling stage	4-modeling stage
Versioned Complex System	6			1	5
Versioned Protocol	6	1		4	1
Versioned Industrial Workshop	23	13	4	3	3
Theoretical	7	4		2	1

For examining whether all the models can be considered together rather than as four distinct groups, we initiated the non-parametric Kruskal-Wallis test. Comparing the six model information measurers between the four model groups revealed no significant differences in all six measurers that related to the last model sample.

We assumed that the six model information measurers should change over time. We initiated a repeated measures ANOVA test to examine this assumption. The two within independent variables were time (first model version; last model version) and information aspect (Number of OPL sentences in each modeling stage, Specification Patterns WINF per sentence, TWINF per sentence, Ratio of objects to processes, Ratio of environmental things to systematic things, Ratio of physical things to informatical things). The dependent variable was the model grade in each measure. In line with our hypothesis, we found a significant difference between the first model versions and the last model versions in all the information aspects, $F(1, 23)=31.39$, $p<.001$, $\eta^2=.57$. We initiated post-hoc related samples t-test to examine each one of the six measurers separately. Table 2 presents the Means and the paired-sample t-test results for each measure. The results indicate significant differences in all the six information measurers.

Table 2 – Comparison of the first and last modeling stage scores for the six infomration measures

Information measure	First modeling stage average	Last modeling stage average	t(23)
Number of OPL sentences in each modeling stage	194.13	308.75	5.69***
Specification Patterns WINF per sentence.	0.32	0.34	2.9**
TWINF per sentence	0.29	0.31	3.4**
Ratio of objects to processes	0.62	0.58	2.38*
Ratio of environmental things to systematic things	0.14	0.11	2.06*
Ratio of physical things to informatical things	0.44	0.35	3.4**

$$*p<0.05;**p<0.01;***<0.001$$

4. Discussion

Based on our analysis and findings, we have drawn the following conclusions about the evolution of conceptual models in general and about model informativity in particular:

A. From each modeling stage to the next, the model grows bigger, i.e., more information is added to the model, and the number of OPL sentences increases.

B. The information added to the model in each stage elaborates the information from the previous stage, details in at least in one more depth level, covers different scenarios, etc. The added information is more informative, detailed, rich, or semantically important.

C. The Specification Patterns informativity increases, while the model specification uncertainty decreases from stage to stage when the information from the previous stage remains in the model for the next stages, and remodeling (which could increase uncertainty) is minimal.

D. The ratio of objects to processes decreases as the modeling progresses, and the emphasis is transferred from objects to processes.

E. The ratio of system things to environmental things in the model increases as the modeling progresses, and increasingly more emphasis is placed on systemic things.

F. The ratio of physical things to informatical things decreases as the modeling progresses. This might have to do with the models being biased towards information systems, in which physical things are usually initially related to the environment, and later to the physical means, such as hardware, endpoint devices, and the communication infrastructure. However, the proportion of informatical things is still prevalent.

5. FUTURE RESEARCH

We propose to study the impact of MIL on the modeling process in the following ways:

A. Provide teams with their informativity scores, so they can act upon this information.

B. Determine how communicating informativity to the analyst encourages/affects direct investments in informativity-lacking areas of the model.

C. Study the response of engineering teams to their model informativity scorecard. Ultimately, attempt to confirm the hypothesis that MIL evaluation and its reflection improves the quality of the MBSE process and of the model, contributing to MBSE return-on-investment (ROI).

D. Qualitatively evaluate the added value of MBSE to system development, based on reviewing the reports and interviewing key personnel from participating organizations.

E. Search for trends in the changes of MIL components over many modeling stages and reason over the existence or lack of such a trend.

F. Capture the semantics the model expresses and analyze the informativity by examining also the semantics, e.g., does a name of a thing (object or process) in the model informs its priority.

G. Compare MIL to the utility that can be derived from the model in aspects such as communicating the system's architecture to stakeholders, the ability to implement the model, and the ability to use the model for simulation as a black box or as a glass box.

H. Examine the cost (time, effort, and resources), versus the benefit of adding information to the model.

ACKNOWLEDGMENTS

We thank the Berny Gordon Center for Systems Engineering at Technion – Israel Institute of Technology for funding this research.

REFERENCES

[1] D. Dori, “Model-Based Systems Engineering with OPM and SysML,” *Springer*, 2016.

[2] “SYSTEMS ENGINEERING VISION 2020,” 2007.

[3] D. Long and Z. Scott, *A primer for model-based systems engineering*2011 ..

[4] D. Dori, *Object-Process Methodology: A Holistic Systems Approach*. Berlin, Heidelberg, New York: Springer, 2002.

[5] D. Dori, *Model-Based Systems Engineering with OPM and SysML*. New York: Springer, 2016.

[6] ISO/TC 184, “ISO/PDPAS 19450 Automation systems and integration — Object-Process Methodology,” 2015. .

[7] J. Estefan, “Survey of model-based systems engineering (MBSE) methodologies,” *INCOSE MBSE Focus Group, 25(8)*2008 ,.

[8] A. L. Ramos, J. V. Ferreira, and J. Barceló, “Model-based systems engineering: An emerging approach for modern systems,” *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 42, no. 1, pp. 101–111, 2012.

[9] Y. Mordecai and D. Dori, “Model Informativity Analysis: Quantifying the Model Informativity Analysis: Quantifying the Expressive Power of Conceptual system Models,” *26thINCOSE Int. Symp. Orlando FL, USA, April 2016.*, 2016.

[10] Y. Mordecai, “Cyber-Physical Disruption Modeling, Analysis, and Management: an Evolutionary Object-Process Model-Based Robust Systems Engineering Approach,” Ph.D. Research Thesis, Technion – Israel Institute of Technology, Haifa, Israel, February-2016. Advisor: Prof. Dov Dori, 2016.

[11] G. van Rossum, “, Python tutorial, Technical Report CS-R9526, Centrum voor Wiskunde en Informatica (CWI), Amsterdam,” 1995.